



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

Data article

## Glycosylation characterization of therapeutic mAbs by top- and middle-down mass spectrometry

Bao Quoc Tran<sup>a</sup>, Christopher Barton<sup>b</sup>, Jinhua Feng<sup>b</sup>, Aimee Sandjong<sup>b</sup>,  
Sung Hwan Yoon<sup>a</sup>, Shivangi Awasthi<sup>a</sup>, Tao Liang<sup>a</sup>, Mohd M. Khan<sup>a</sup>,  
David P.A. Kilgour<sup>a</sup>, David R. Goodlett<sup>a,\*</sup>, Young Ah Goo<sup>a,\*</sup>

<sup>a</sup> Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD, USA

<sup>b</sup> MedImmune, LLC, Gaithersburg, MD, USA

## ARTICLE INFO

## Article history:

Received 20 October 2015

Received in revised form

4 November 2015

Accepted 10 November 2015

Available online 24 November 2015

## ABSTRACT

A reference monoclonal antibody IgG1 and a fusion IgG protein were analyzed by top- and middle-down mass spectrometry with multiple fragmentation techniques including electron transfer dissociation (ETD) and matrix-assisted laser desorption ionization in-source decay (MALDI-ISD) to investigate heterogeneity of glycosylated protein species. Specifically, glycan structure, sites, relative abundance levels, and termini structural conformation were investigated by use of Fourier transform ion cyclotron resonance (FT-ICR) or high performance liquid chromatography electrospray ionization (HPLC-ESI) linked to an Orbitrap. Incorporating a limited enzymatic digestion by immunoglobulin G-degrading enzyme *Streptococcus pyogenes* (IdeS) with MALDI-ISD analysis extended sequence coverage of the internal region of the proteins without pre-fractionation. The data in this article is associated with the research article published in *Journal of Proteomics* (Tran et al., 2015) [1].

© 2015 Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <http://dx.doi.org/10.1016/j.jprot.2015.10.021>

\* Corresponding authors.

E-mail addresses: [dgoodlett@rx.umaryland.edu](mailto:dgoodlett@rx.umaryland.edu) (D.R. Goodlett), [ygoo@rx.umaryland.edu](mailto:ygoo@rx.umaryland.edu) (Y.A. Goo).

<http://dx.doi.org/10.1016/j.dib.2015.11.031>

2352-3409/© 2015 Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	Biology, Chemistry
More specific subject area	Top- and middle-down mass spectrometric analysis of protein species and proteoforms
Type of data	Table, Figure
How data was acquired	Mass spectrometry data were collected on a 12TFT-ICR XR (Bruker) and a LTQ-Orbitrap Elite (Thermo Scientific)
Data format	MS spectra
Experimental factors	Top- and middle-down analyses with multiple fragmentation techniques including electron transfer dissociation (ETD) and matrix-assisted laser desorption ionization in-source decay (MALDI-ISD) for characterization of a reference monoclonal antibody IgG1 and a fusion IgG protein
Experimental features	Glycan profiling including structure, sites, relative abundance levels, and termini structural conformation were investigated
Data source location	Baltimore and Gaithersburg, Maryland, USA
Data accessibility	Data is provided within this article

## Value of the data

- Application of top-down and middle-down mass spectrometry for characterization of comprehensive glycosylated protein species from recombinant IgG and IgG-fusion proteins.
- Detection of glycan structure, sites, and relative abundance levels.
- Investigation of termini structural conformation; N-terminal glutamine (Q) to pyroglutamate (pyroGlu, aka. pE) conversion and C-terminal lysine truncation.
- Improvement on protein sequencing by MALDI ISD, ETD in combination with middle-down analysis by use of an enzyme, IdeS.

## 1. Data

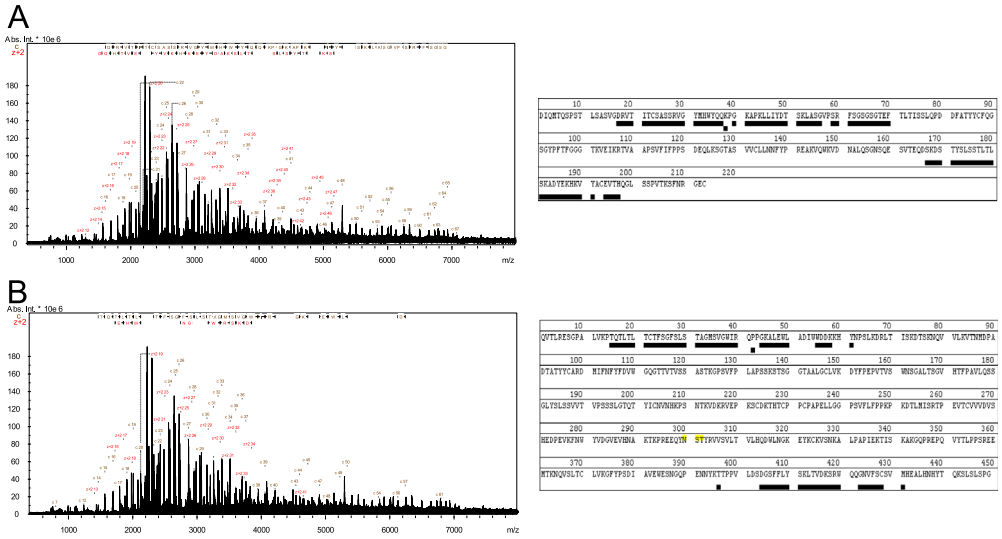
Glycosylation profiling of recombinant IgG and IgG-fusion proteins and sequence coverage information.

## 2. Experimental design, materials and methods

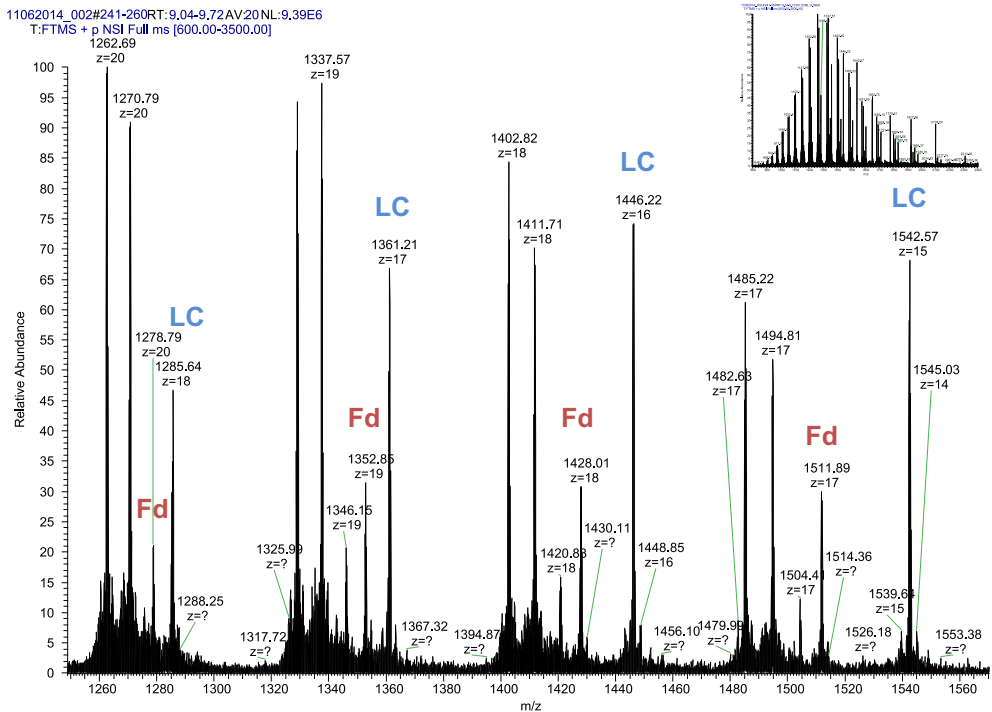
The recombinant human mAb (IgG1) and IgG1-fusion protein were manufactured using standard cell culture, purification and formulation processes [2]. The IgG1 has two identical light chains and two identical heavy chains with molecular weight of 150 kDa and N-glycosylation site at N297. The IgG fusion protein, approximately 90 kDa, was comprised of two chains each with a human protein domain fused onto the N-terminus of a human IgG1 constant domain (Fc). The protein has three N-glycosylation sites at N76, N108 and N207.

## 3. Analysis of the IgG protein

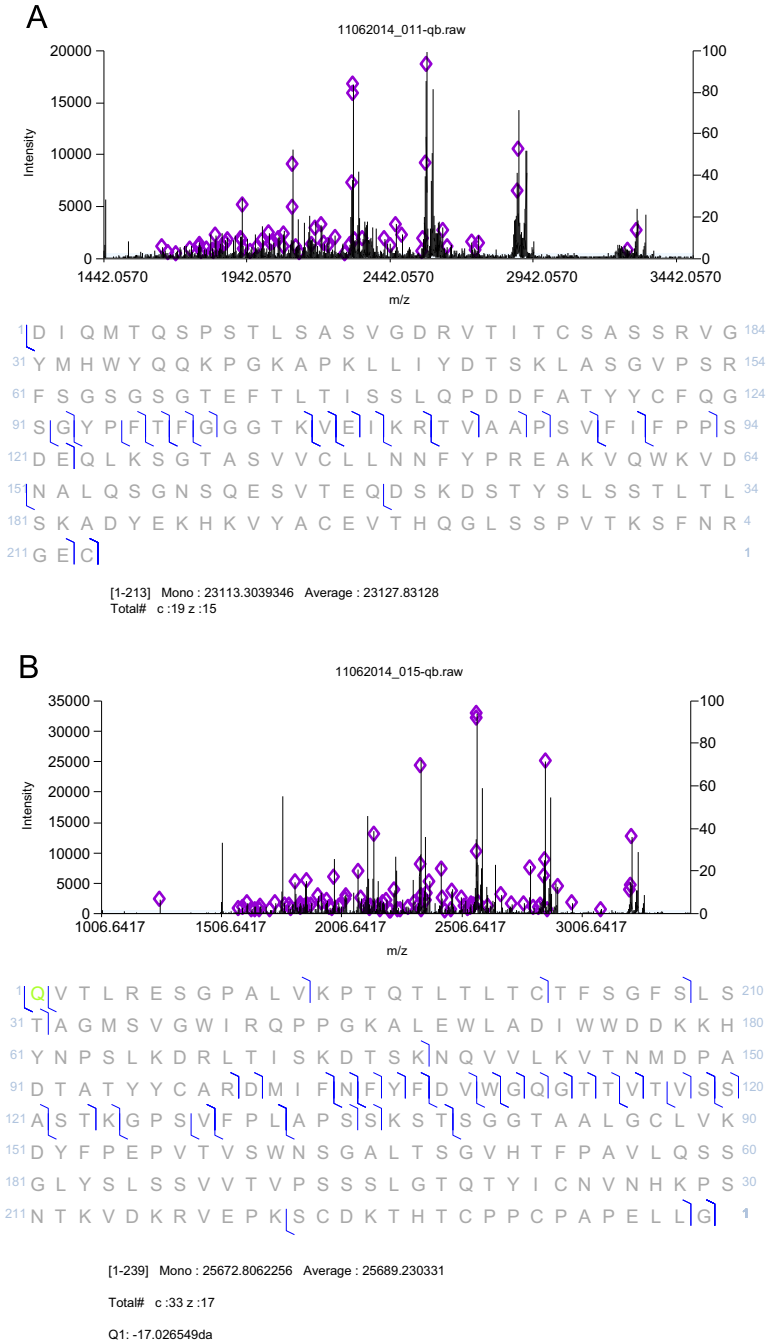
The IgG protein was analyzed 1) in reduced form by MALDI-ISD in a 12 T Fourier Transform Ion Cyclotron Resonance (FT-ICR XR) mass spectrometer (Bruker, Bremen, Germany), and 2) after IdeS digestion and dithiothreitol (DTT) reduction to generate three smaller fragments; light chain, Fab, and Fc/2 after which analyzed by LC-MS/MS on an Orbitrap Elite (Thermo Scientific, Saint Jose, CA, USA).



**Fig. 1.** MALDI-MS/MS mass spectrum of the DTT-reduced IgG1 mAb. Amino acid residues underlined were identified after phase correction by c or z + 2 ions with a mass error threshold of 50 ppm, covering (A) 37.6% of the light chain and (B) 14.3% of the heavy chain.



**Fig. 2.** LC-MS/MS mass spectrum of IdeS/DDT treated IgG mAb. The charge state distribution of light chain (LC) and Fd domain are shown. Inset shows the full mass spectrum collected after accumulation of 20 scans over an  $m/z$  range 600–3500 at a resolving power of 120,000.



**Fig. 3.** (A) ETD mass spectrum of precursor  $m/z$  1446 (charge 16+) was searched against IgG1 mAb light chain sequence with mass tolerance of 100 ppm. ETD mass spectrum was produced from accumulation of 49 scans. Identified c- and z- fragment ion series are indicated in diamonds in the spectrum and relevant sequence identification is shown in the sequence. (B) ETD mass spectrum of precursor  $m/z$  1512 (charge 17+) was searched against Fd sequence of IgG1 mAb with mass tolerance of 100 ppm. ETD mass spectrum was produced from accumulation of 50 scans. Identified c- and z- fragment ion series are indicated in diamonds in the spectrum and relevant sequence identification is shown in the sequence. N-terminal glutamine (Q) to pyroglutamate (pyroGlu) conversion was confirmed with precursor mass error 3.9 ppm (0.10 Da).

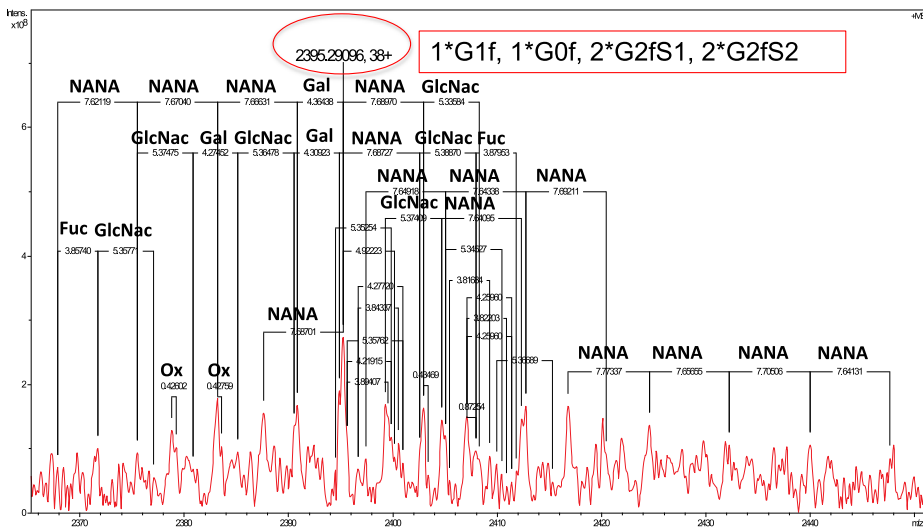
Intact protein mass measurement for detection of protein species of the IgG was performed using ESI FT-ICR MS as demonstrated in the related study [1].

### 3.1. MALDI-ISD analysis of reduced protein

The IgG protein in 1  $\mu\text{g}/\mu\text{L}$  in 50 mM ammonium bicarbonate buffer was reduced to light and heavy chain by 20 mM dithiothreitol (DTT, product 43815, Sigma-Aldrich, St. Louis, MO, USA) at 80 °C for 15 min. The reduced protein sample was buffer exchanged to 5% acetonitrile (ACN) /0.1% formic acid (FA) using a 10 kDa molecular weight cut off (MWCO) centrifugal filter (product UFC201024, Merck Millipore, Tullagreen, Carrigtwohill, Ireland) and then diluted to 17  $\mu\text{g}/\mu\text{L}$  in 50% ACN/0.1% FA. 1  $\mu\text{L}$  sample and 1  $\mu\text{L}$  of 1,5-diaminonaphthalene (1,5-DAN, product 56451, Fluka, St. Louis, MO, USA) saturated in ACN were spotted on a stainless steel MALDI plate. MALDI-ISD experiment was performed with a transient domain of 1 mega-word data points (estimated resolving power of 270,000 at  $m/z$  400) covering  $m/z$  range 400–10,000. MALDI-ISD mass spectra after phase correction [3] were processed to identify  $c$ - and  $z+2$  ions from light chain and heavy chain with mass tolerance 50 ppm using DataAnalysis 4.2 and Biotools software 3.2 (Bruker) (Fig. 1).

### 3.2. LC-MS/MS analysis of IdeS-digested protein

The IgG protein was digested with IdeS (product A0-FR1-020, Genovis AB, Lund, Sweden) at a ratio of 1 unit of enzyme per 1  $\mu\text{g}$  protein for 2 h at 37 °C, followed by DTT treatment. Approximately 14  $\mu\text{g}$  of the IdeS/DTT treated sample was separated on a 100  $\mu\text{m} \times 20$  mm column packed with Jupiter 5  $\mu\text{m}$  C5 particles (product 04A-4052, Phenomenex, Torrance, CA, USA) by a fast mobile gradient of 5–50% ACN/0.1% FA in 15 min at a flow rate of 0.3  $\mu\text{L}/\text{min}$ . Light chain and Fd domain were detected in a range of 350–3500  $m/z$  at resolving power of 120,000 in positive ion mode on the Orbitrap Elite (Fig. 2). Precursor masses were selected for ETD fragmentation. ETD spectra were acquired with 5 microscans at resolving power of 120,000 and processed by MASH Suite version 1.0 software [4]. The fragment ions and corresponding sequence coverage are shown in Fig. 3A for the light chain and Fig. 3B for the Fd domain. Glutamine (Q) to pyroglutamate (pyroGlu) conversion was confirmed on the N-terminal of heavy chain (Fig. 3B).



**Fig. 4.** Fourier transform ion cyclotron resonance (FT-ICR) mass spectrum of the intact fusion protein peaks (charge 38+). The mass spectrum was produced from accumulation of 1650 scans collected in a time domain of 128 K words data points. Major peaks demonstrate a heterogeneous combination of glycosylation, predominantly by sialic acids (NANA). Additional information regarding the glycan nomenclature and structure are provided in Fig. 5.

#### 4. Analysis of the IgG fusion protein

The fusion protein was analyzed 1) by accurate mass measured in FT-ICR for protein species detection 2) by MALDI-ISD following DTT reduction in FT-ICR, and 3) by accurate mass measurement following IdeS cleavage in FT-ICR or ETD MS/MS in the Orbitrap Elite.

##### 4.1. Accurate mass measurement of intact protein

The fusion protein sample was buffer exchanged to 5%ACN and 0.1%TFA using Amicon 50 kDa MWCO centrifugal filters (product UFC505024, Merck Millipore). Since the fusion protein sample was stored in buffer with detergent, detergent removal was performed by Pierce detergent removal kit (product 87778, Thermo Scientific, Rockford, CA, USA). The detergent-free protein sample was diluted to 10  $\mu\text{g}/\mu\text{L}$  final concentration in 50% ACN/0.1% FA and introduced to electrospray ionization (ESI)-FT-ICR for intact protein mass measurement for detection of protein species. The experiment was carried out in positive ion mode on the 12 T FT-ICR at flow rate 2  $\mu\text{L}/\text{min}$ . Mass spectra were collected in magnitude mode with a transient domain of 128,000 word data points to cover  $m/z$  range 400–4000, with an accumulation of 1650 scans. Fig. 4 shows the protein species detected in  $m/z$  region 2365–2450 with high heterogeneity of glycosylation. Corresponding glycan nomenclature and structures are shown in Fig. 5.

##### 4.2. MALDI-ISD analysis of reduced protein

The protein sample was diluted to 1  $\mu\text{g}/\mu\text{L}$  in 50 mM ammonium bicarbonate buffer and reduced with 20 mM DTT at 80 °C for 15 min. The reduced sample was then detergent removed. The

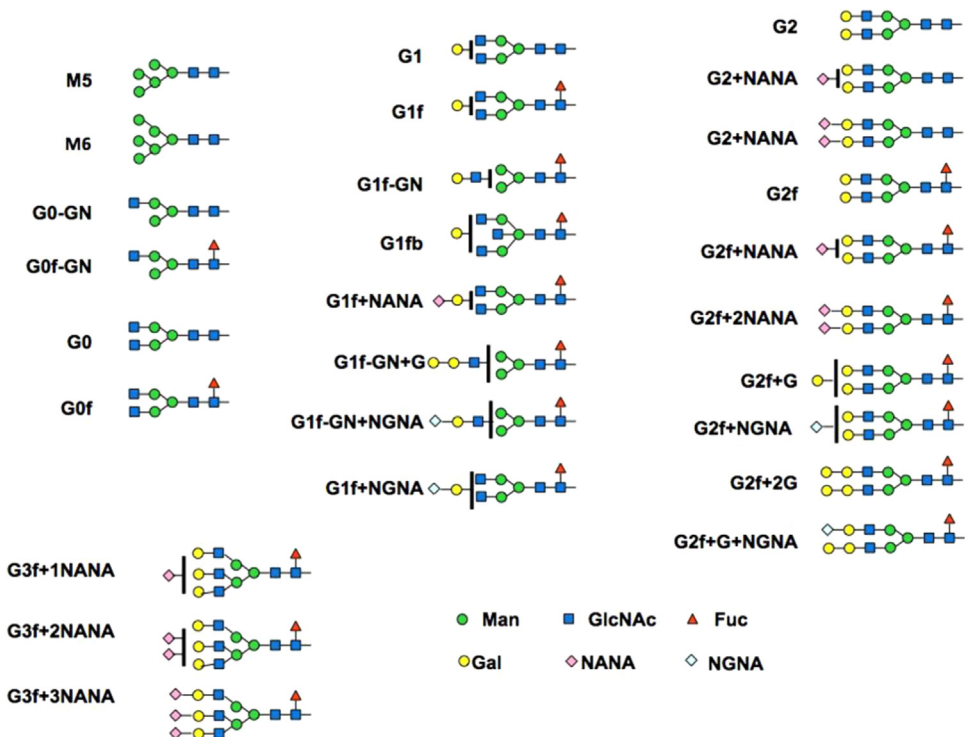
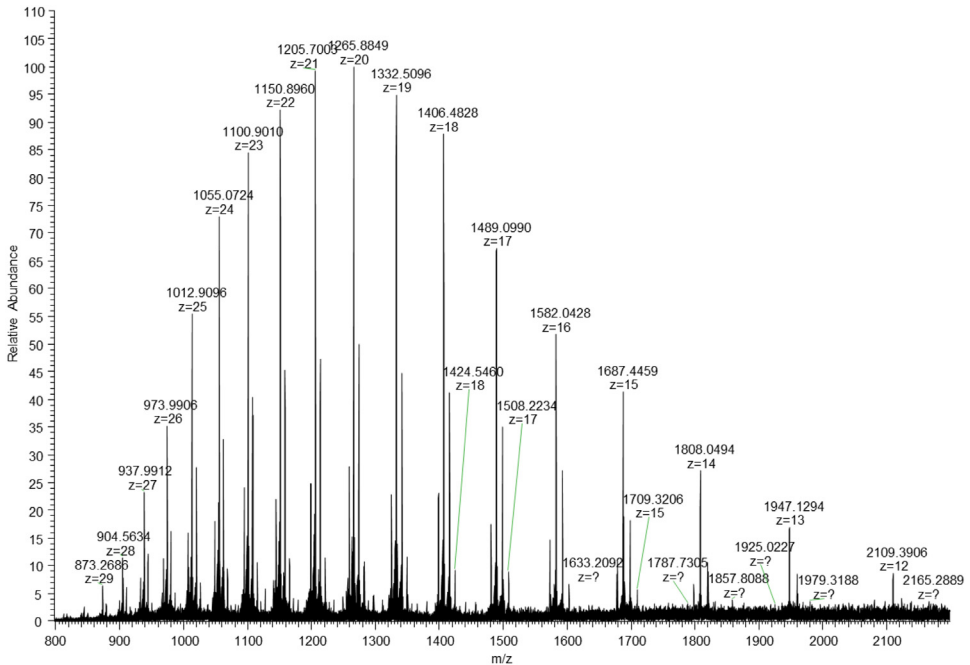


Fig. 5. The glycan nomenclature and structure.





**Fig. 7.** LC-MS mass spectrum of IdeS/DDT treated fusion protein. Charge state distribution of glycosylated protein species from of C-terminal fragment is shown. The mass spectrum was produced from accumulation of 20 scans over an  $m/z$  range 600–3500 at a resolving power of 120,000.

#### 4.3. ETD MS/MS analysis of IdeS-digested protein

The fusion protein sample was diluted to  $1 \mu\text{g}/\mu\text{L}$  in 50 mM ammonium bicarbonate solution and digested with enzyme IdeS at ratio of 1 enzyme unit to  $1 \mu\text{g}$  protein for 2 h at  $37^\circ\text{C}$ . 20 mM DDT was added to cleave disulfide bonds at  $80^\circ\text{C}$  for 15 min. The sample was further concentrated using 10 kDa MWCO filter and subject to detergent removal. The sample buffer was exchanged to 5% ACN/0.1% FA and diluted to final concentration of approximately  $7 \mu\text{g}/\mu\text{L}$  in 50%ACN/0.1% FA. IdeS digestion and DTT reduction cut the fusion protein to half the size producing N-terminal and C-terminal fragments.

Accurate mass measurement of the IdeS/DDT treated sample was performed by ESI FT-ICR. Mass spectra were collected in magnitude mode with a transient domain of one mega-word data points to cover  $m/z$  range 400–4000. Seven glycoforms were identified from the C-terminus at multiple charge states. Table 1 represents glycosylated protein species detected from 17+ charge peaks and their glycan nomenclature and structures are shown in Fig. 5.

The sample was also directly infused into the Orbitrap Elite. Data were collected in positive ion mode for 350–3500  $m/z$  range at resolving power of 120,000 (Fig. 7). Precursor masses of the glycoforms were isolated for ETD fragmentation. ETD mass spectra were acquired with 5 microscans at resolving power of 120,000 and processed by MASH Suite for sequence coverage analysis. Manual interpretation of the ETD spectra confirmed that glycosylation by G0f or G1f occurred on a site N207 and the protein has a C-terminal lysine truncation (Fig. 8).



